

Artificial Intelligence Identification of Japonica Rice varieties Based on Raman Spectroscopic Identification Mechanism of Saline-alkali Tolerance

Rui Liu

Qiqihar University, Qiqihar 161006

Bo Ma

Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161006

Xin Zhao

Qiqihar University, Qiqihar 161006

Kefei Tan

Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161006

Jifang Hu

Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161006

Ming Yuan

Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161006

Lianxia Wang

Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161006

Hongbo Xiang

Heilongjiang University of Technology, Jixi 158100

Dongxue Mi

Qiqihar Institute of Engineering, Qiqihar 161006

Yongcai Lai

game0451@163.com

Northeast Branch of National Saline-Alkali-Tolerant Rice Technology Innovation Center

Kai Liu

Northeast Branch of National Saline-Alkali-Tolerant Rice Technology Innovation Center

Shibo Wang

Qiqihar University, Qiqihar 161006

Article

Keywords: Raman spectrum, Artificial Intelligence, Japonica rice, Saline-alkali tolerant, Identification mechanism

Posted Date: September 9th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4904829/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Artificial Intelligence Identification of Japonica Rice varieties Based on Raman Spectroscopic Identification Mechanism of Saline-alkali Tolerance

Bo Ma ^{1,2,†}, Rui Liu ^{3,4,†}, Dongxue Mi ⁵, Hongbo Xiang ⁶, Shibo Wang ^{3,4}, Ming Yuan ¹ and Lianxia Wang ¹, Jifang Hu ¹, Kefei Tan ¹, Yongcai Lai ^{2,*}, Kai Liu ^{2,*} and Xin Zhao^{3,*}

¹ Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161006, China;mabo8210@haas.cn (B.M.); y.m@haas.cn (M.Y.); wlx0427@haas.cn (L.W.)
hujifang7@haas.cn (J.H.); tkfhj@haas.cn (K.T.)

² Northeast Branch of National Saline-Alkali-Tolerant Rice Technology Innovation Center, Harbin 150000, China;mabo8210@haas.cn (B.M.); yame0451@163.com (Y.L.); liukailouis@163.com (K.L.)

³ Qiqihar University, Qiqihar 161006, China;zxsnh@hrbeu.edu.cn (X.Z.);411055273@qq.com(R.L.);wangshibo05@163.com(S.W.)

⁴ Nenjiang River Basin Regional Innovation and Industrial Integration Development Laboratory, Qiqihar, Heilongjiang 161006, China;411055273@qq.com(R.L.); wangshibo05@163.com(S.W.)

⁵ Qiqihar Institute of Engineering, Qiqihar 161006, China;18845281616@163.com

⁶ Heilongjiang University of Technology, Jixi 158100, China;xhb_2005@163.com (H.X.)

* Correspondence: yame0451@163.com (Y.L.); liukailouis@163.com (K.L.);zxsnh@hrbeu.edu.cn (X.Z.)

† These authors contributed equally to this work.

Abstract: Rice is regarded as the preferred crop for saline-alkali soil improvement by researchers. At present, the identification method for saline-alkali tolerance of rice varieties requires researcher to conduct tedious field investigations based on growth indicators. Therefore, there is an urgent need for an effective technical means to quickly and accurately identify saline-alkali tolerance of rice varieties. Study used 20 japonica rice varieties with three types of saline-alkali tolerance as test materials, by analyzing the identification mechanism of salt-alkali tolerance in Raman spectrum of japonica rice varieties, seven characteristic spectral peaks closely related to salt-alkali tolerance were identified. Various algorithms in Python are used for data standardization, baseline elimination, extraction of characteristic spectral peaks, detection of peaks characteristic information and data noise reduction. Three identification models were established to confirm the highest accuracy of CapsNets identification model, which could provide technical support and reference for breeding saline-alkali resistant japonica rice varieties.

Keywords: Raman spectrum; Artificial Intelligence; Japonica rice; Saline-alkali tolerant; Identification mechanism

1.Introduction

Soil salinization can reduce crop yield to varying degrees, and can lead to the reduction of cultivated land area in severe cases^[1]. As a moderately saline-alkali tolerant aquatic crop, rice is regarded by researchers as the preferred crop for saline-alkali soil improvement by taking advantage of "washing salt and alkali with water and removing salt and alkali with water"^[2]. The total saline-alkali land area of Heilongjiang Province in China is 1.34 million hm², and there are about 1 million hm² saline-alkali land rich in water resources^[3]. Japonica rice planted in saline-alkali land rich in water resources can dissolve and discharge soluble salt and alkali in

saline-alkali soil through irrigation, and japonica rice roots can absorb salt and secrete organic acids, thus improving soil saline-alkali conversion. Reduce soil saline-alkali damage and increase cultivated land resources^[4,5]. It can be seen that japonica rice plays an irreplaceable role in the improvement of saline-alkali land. However, the saline-alkali tolerance of different japonica rice varieties varies greatly, so rice breeders attach great importance to the identification and evaluation of saline-alkali tolerance japonica rice varieties.

Traditional screening and breeding of saline-alkali tolerant japonica rice^[6-9] mainly rely on the investigation and analysis of phenotypic indexes and the method of molecular marker-assisted selection^[10-12]. However, this method requires traditional phenotypic index data and skilled operation of professionals with advanced equipment to analyze complex genetic data, resulting in complex detection process and high cost. Therefore, a rapid and accurate identification technique and method are urgently needed to evaluate the saline-alkali tolerance of japonica rice varieties.

Raman spectroscopy, based on the Raman scattering effect discovered by Indian physicist C.V. Raman^[13,14], is a technique for studying the structure and chemical properties of a substance by measuring the vibration patterns of its molecules or atoms at specific frequencies. It has the advantages of fast, efficient, accurate, non-destructive, high sensitivity and good stability, so it is widely used in scientific research and practical applications^[15-17].

Python is a popular programming language that is concise and easy to learn and use^[18]. It not only supports high-level language functions such as matrix manipulation, but also can be applied to Web development, data analysis, image recognition and system construction. In machine learning methods, Python can visualize data from different fields and further analyze the visualization results^[19-21]. Python can also effectively process spectral data according to its data visualization characteristics, and realize functions such as spectral data preprocessing, spectral peak extraction, spectral peak feature selection, cluster analysis, classification and identification, and image recognition^[2,22].

In view of this, this study took japonica rice varieties with different salt-alkali tolerance degrees in Heilongjiang Province, China as the object, proposed a salt-alkali tolerance japonica rice identification model based on Raman spectroscopy and combined with machine learning algorithm in Python, and analyzed its salt-alkali tolerance identification mechanism by Raman spectroscopy. Rapid and accurate identification of saline-alkali tolerance japonica rice was realized, which provided a new idea and method for breeding saline-alkali tolerance japonica rice varieties efficiently. In summary, this study has important theoretical and application value, and is of great significance for promoting saline-alkali tolerance japonica rice variety breeding and saline-alkali land improvement and utilization.

2. Materials and Methods

2.1. Test Materials

Twenty japonica rice varieties (lines) for salt and alkali tolerance identification were provided by Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences in 2022, numbered according to S1 to S20, and were suitable for planting in Heilongjiang Province. The experiment was conducted in the Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences in 2022 (Table 1). The treated soil was soda saline-alkali soil with a pH of 9.2 and a salt content of 0.41%, while the control soil was common paddy soil with a pH of 7.1 and a salt content of 0.18%.

Table 1. Information of japonica rice varieties and sources for testing.

Number	Varieties (lines)	Provenance	Plants date	Sampling date
S1	KD10	Qiqihar City in Heilongjiang Province, China	April 2022	October 2022
S2	KD15			
S3	SJ6			
S4	LJ9			
S5	LJ12			
S6	LJ21			
S7	LJ39			
S8	LH05-379			
S9	SJ2			
S10	SJ10			
S11	QJ8			
S12	QJ28 (22)			
S13	DN416			
S14	DN9220			
S15	LD10			
S16	LD12			
S17	SJ3			
S18	LQD3			
S19	JD58			
S20	FY305			

2.2. Test Methods

2.2.1. STC Method for Screening Saline-alkali Tolerance of Japonica Rice Varieties

The saline-alkali tolerant coefficients(STC) of rice , that is, the relative growth value of each salt and alkali tolerance index (Formula 1), was evaluated and classified to eliminate the differences in basic traits among the tested materials [23].

$$\text{STC} = \text{Saline-alkali stress index} / \text{Control index value} \quad (1)$$

Note: The saline-alkali stress index value is one of the agronomic traits index value in the treated field, the control index is the index value of one of the agronomic traits in the control field.

2.2.2. Acquisition method of Raman Spectrum of Saline-alkali Resistant Japonica Rice Varieties

The test temperature was about 23°C and the relative humidity was 57%RH. Japonica rice grains were husked by Shanghai LJJM rice cleaning machine, and each variety sample was husked for 50's and treated to remove plant fluorescent substances, and put into an envelope to be tested. The data information of the sample to be tested is collected by the benchtop Raman spectrometer (model: Advantage 532) produced by DeltaNu, which adopts 5mW532nm excitation and CCD refrigeration to achieve high Raman excitation power and high SNR. The full spectrum measurement range is 0-3400cm⁻¹, and the primary spectrograph range (effective measurement range) of the sample to be measured in this test is 400cm⁻¹-3400cm⁻¹, with a resolution of 10cm⁻¹. Set the parameters of the Advantage 532 and NuSpecTM software by adjusting the height and position of the sample holder (Table 2). Using NUSpecTM software to store the two-dimensional data information of samples and observe the original two-dimensional image information of samples, the obtained spectral data information is saved in prn format file.

Table 2. Parameter Settings for Advantage 532 and NuSpec™

Laser Power	Stimulate Wavelength	Integration Time	Number of Spectrum	Display	Save Spectrum	Resolution
High	532nm	4	3	Average	ASCII	Low

2.2.3.Acquisition Method of Raman Spectrum Characteristic Peak of Japonica Rice

(1) Standardization of spectral data

NL normalizes the spectral data to a certain range, and can adjust the spectral intensity to a relative or relatively comparative scale, making the comparison between different samples more accurate, so as to better compare and analyze the spectral data. Z-score normalization is adopted, by subtracting the mean value of the spectrum, and then dividing by the standard deviation of the spectrum, so that the mean value of the spectrum is 0 and the standard deviation is 1, as shown in formula 2:

$$Z - score : z = \frac{x - u}{\delta} \quad (2)$$

Note: z is the normalized value, x is the raw data; u is the mean of the data, δ is the standard deviation of the data.

(2) Baseline elimination of spectral data

CWT has the capability of localized feature analysis in both time and frequency domains, and can translate forward and backward along the time axis, and stretch and compress proportionally to obtain low and high frequency wavelets [24,25]. In order to extract useful signals more efficiently and filter out interfering signals, a specific wavelet function (formula 3) is constructed.

$$\psi_{a,b(t)} = \frac{1}{\sqrt{|a|}} \psi \left[\frac{t-b}{a} \right], a, b \in R, a \neq 0 \quad (3)$$

Experiment with calling the signal.cwt (x, signal.ricker, widths) statement to perform a continuous wavelet transform: deep copy the input data and assign it to x, using Ricker wavelet as mother wavelet, widths is an array containing integers from 1 to w-1, where w is the window size of the wavelet transform.

(3) The method of obtaining the characteristic spectral peak

In this experiment, the method scipy.signal.find_peaks (SSFP)[2,22] in Python was used to extract the Raman spectral peaks of japonica rice. The height parameter in SSFP was used to extract the significant spectral peak of the Raman spectral characteristics of japonica rice varieties. The prominence, width, and dif parameters were selected to capture the prominence, width, and offset information.

2.2.4.The Method of Raman Spectroscopic Identification Mechanism of Salt-alkaline Tolerant Japonica Rice

Amino acid is the basic unit of protein. Through the analysis of biological mechanism, amino acid plays an important regulatory role in saline-alkali tolerant japonica rice. Amino acids are organic compounds consisting of an amino group (-NH₂), a carboxyl group (-COOH), and a side

chain (-R). Different amino acids have different side chain structures, and the side chain of an amino acid refers to the part of the amino acid molecule that is connected to the alpha-carbon atom, also known as the R group (R-group). These side chain structures determine the chemical properties and functions of amino acids. The functional groups in the amino acid side chain exhibit specific vibration patterns through Raman spectroscopy.

2.2.5. The Pretreatment Method of Raman Spectra Data

Spectral smoothing is a common spectral data preprocessing method, which is based on sliding window or filter algorithm to smooth the spectral data, which can eliminate noise and fluctuation and make the spectral curve smoother. In this study, the zero-phase Scipy.signal.filtfilt (Filtfilt) [26] spectral smoothing method was used to preprocess Raman spectral data. Filtfilt applies a linear digital filter twice, with twice the order of the other filters, once forward and once backward, with twice the zero phase and twice the filter order, thus effectively avoiding phase difference. As shown in formulas 4, 5 and 6:

$$y_1 = \text{filtfilt}(b, a, x) \quad (4)$$

$$y_2 = \text{filtfilt}(b, a, \text{reverse}(y_1)) \quad (5)$$

$$y = \text{reverse}(y_2) \quad (6)$$

Note: M is the length of b (molecular coefficient) minus one, N is the length of a (denominator coefficient) minus one, x[n-k] is the value of the input signal x at time n-k, y[n-k] is the value of the output signal y at time n-k.

First, the input signal x is filtered forward once through the filter to obtain the intermediate signal y₁. Then reverse y₁ (that is, reverse in time), and reverse filter through the same filter to get the intermediate signal y₂ after reversal. Finally, y₂ is reversed back to get the final filtering result y.

2.2.6. CapsNets Identification Method

(1) Establishment of logistic regression (LR) Identification Model

LR trains the model on a given set of n data (the training set), and the model learns the intrinsic laws and features of the data in order to classify new, unseen data (the test set). The input function of LR is the result of linear regression, and its prediction function is Sigmoid, as shown in formulas 7 and 8:

$$h(w) = w_1x_1 + w_2x_2 + w_3x_3 \dots + b \quad (7)$$

$$g(\theta^T X) = \frac{1}{1 + e^{-\theta^T X}} \quad (8)$$

Note: w₁, w₂, ..., w_n is the weight, x₁, x₂, ..., x_n is the input feature, b is the offset term, x is the input eigenvector, $\theta^T x$ is the weight vector, $g(\theta^T x)$ is the transpose of the weight vector; The Sigmoid function outputs results in the interval [0,1], where the default machine threshold is 0.5.

(2) Establishment of support vector machine (SVM) identification model

The Kernel list parameters of the SVM model in the research usually use Radial Basis Function(RBF), Linear Kernel(LK), Sigmoid and Polynomial Kernel(PK), and then the algorithm

characteristics of these four kernel functions are analyzed to construct SVM discrimination model.

RBF is a local kernel function, and as the data point becomes farther from the center point, the value becomes smaller (formula 9). RBF has A good anti-interference ability to the noise existing in the data. Due to its strong locality, its parameters determine the function range and weaken with the increase of parameter δ^2 .

$$\text{RBF:} \quad \exp\left(-\frac{1}{2\delta^2}\|X - X_i\|^2\right) \quad (9)$$

LK is the simplest kernel function and is a special case of the radial basis kernel function (formula 10). LK performs well mainly for linearly separable data sets, and because the linear kernel function directly computes the inner product of the two input eigenvectors x and z , it does not involve complex transformations, and the computation is very fast.

$$\text{LK:} \quad K(x, z) = \langle x, z \rangle \quad (10)$$

Sigmoid is derived from neural networks and is widely used in deep learning and machine learning (formula 11). Sigmoid is a commonly used kernel function where x and z are the input data and $\tanh$ is the hyperbolic tangent function, and their dot product reflects the similarity between the data. When the two data points are similar, the dot product is larger and the Sigmoid value is larger. When two data points are not similar, the dot product is small and the value of Sigmoid is also small.

$$\text{Sigmoid:} \quad K(x, z) = \tanh(\eta \langle x, z \rangle + c) \quad (11)$$

Multinomial kernel function PK is suitable for orthogonal normalization (vector orthogonal and modulus 1) data (formula 12). Multinomial kernel function PK is a global kernel function, allowing data points that are far apart to have an effect on the value of the kernel function. The larger the parameter d , the higher the dimension of the mapping, and the greater the computation. When d is too large, it is easy to overfit because the learning complexity is too high.

$$\text{PK:} \quad K(x, z) = (\gamma \langle x, z \rangle + c)^d \quad (12)$$

(3) Establishment of CapsNets Identification Model

According to the characteristics of Raman spectral information of saline-alkali tolerant japonica rice varieties, the shape characteristic information of 7 selected spectral peaks was converted from 2D data information to 3D image information by means of one time intercept, one time segmentation and one time fusion [27]. In this intercept, based on the research results of the identification mechanism of saline-alkali tolerance of different japonica rice varieties by Raman spectroscopy, the spectral peaks at the Raman shift of 480cm^{-1} , 865cm^{-1} , 94cm^{-1} , 1129cm^{-1} , 1339cm^{-1} , 146cm^{-1} and 2910cm^{-1} were selected, and the Raman shift was taken as the center point. Based on each central point, 56 and 55 Raman shift lengths were intercepted forward and backward respectively, and 112 Raman shift lengths were captured for each spectral peak. For each characteristic spectral peak (112 Raman shift) obtained after one intercept, the 28, 56 and 84 Raman shifts of each characteristic spectral peak were segmented, and each spectral peak was segmented into 4 segments. Finally, based on the interception and segmentation of the spectral peak shape feature information, the spectral peak shape feature information is fused, and the spectral peak selected by each sample is sorted according to the Raman shift value of the spectral

peak from small to large. According to the order of the beginning, highest and end of the spectral peak, these pictures are fused, and each sample finally gets a three-dimensional image information.

According to the excellent performance of CapsNets in processing image recognition tasks, the identification of saline-alkali tolerant japonica rice varieties CapsNets based on 3D image information of characteristic spectral peaks was studied. The Capsule in CapsNets is a core component, and its working principle involves many aspects such as vector representation, dynamic routing algorithm and nonlinear transformation [28,29]. As shown in Figure 1, CNN outputs a single scalar value as a result through artificial neurons. This design has certain limitations when processing images, and some key spatial information will be lost, resulting in the inability to effectively process the relative position relationship between features. In contrast, CapsNets uses capsules as the basic unit, and each capsule can output a vector to represent the result, which not only contains the existence probability of the entity, but also encodes other important attributes of the entity, such as posture, position and possible deformation, etc., the length of the vector represents the probability of the existence of the entity, and the direction of the vector contains other attribute information of the entity. The connection between capsules is realized by dynamic routing algorithm. The algorithm allows information to flow between capsules and dynamically adjusts the weights of the connections based on the input data and the output of the current network. This dynamic adjustment mechanism allows CapsNets to better adapt to changes in input data and learn more complex feature representations. In addition, a nonlinear transformation was carried out inside the capsule to enhance the expressive power of the network. This nonlinear transformation is usually achieved through a fully connected layer or activation function, which is capable of nonlinear mapping of input vectors to capture more complex features.

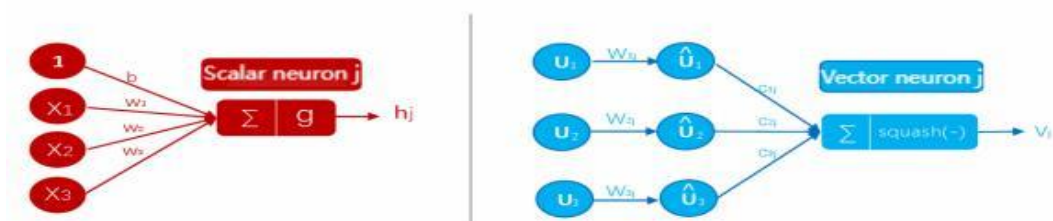


Figure 1. Schematic diagram of scalar and vector output

CapsLayer implementations in CapsNets are typically divided into three layers, the first being the initial Layer (Conv1), the second being the Primary Capsule Layer, and the third being the Final CapsLayer. Figure 2 shows that Conv1 is a standard Convolutional Layer, which is responsible for extracting local features of the input image, and uses the Rectified Linear Unit (ReLU) activation function to convert negative values to 0, thus introducing nonlinearity and enhancing the expression capability of the network. The low-level feature map obtained after processing the Conv1 layer is used as input to the main capsule layer. The primary capsule layer is responsible for further processing the low-level feature map output from the initial layer into the form of capsules, using a nonlinear compression function to ensure that the vector length of each capsule is between 0 and 1, thus conforming to the characteristics of the probability distribution. Final CapsLayer is responsible for further processing the output information of the main capsule layer into higher-level capsules, whose output is a set of vectors representing the existence probability of a specific category and its related attributes (posture and position), ensuring that

capsules of each category can fully take into account all relevant information in the input image. To make more accurate judgments.

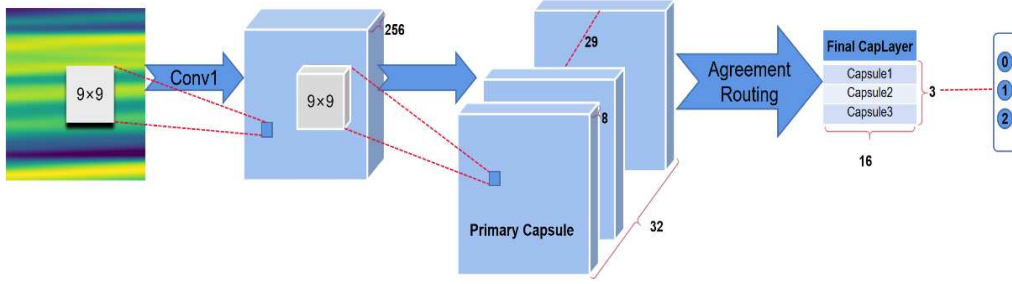


Figure 2. Schematic diagram of the CapsLayer model

Using the nonlinear compression function (formulas 5-13) to compute the output capsule, this transformation is actually a compression and unitization of the vector s , if s is long, then v is approximately equal to 1, if s is short, then v is approximately equal to 0. The length of the output v vector is between 0 and 1, so the length can be expressed as a characteristic probability value.

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (5-13)$$

Note: v_j is the vector output of capsule j , s_j (formula 5-14) is the input vector of capsule j , where $\|s_j\|$ is the length of the s vector.

$$s_j = \sum_i c_{ij} \hat{u}_{j|i} \quad (5-14)$$

Note: c_{ij} (formula 5-15) is the coupling coefficient, $\hat{u}_{j|i}$ (formula 5-16) is the output prediction vector of capsule j at a higher level calculated from capsule i at the next layer.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (5-15)$$

Note: c_{ij} is the logarithmic probability that capsule i will couple to capsule j and is initially set to 0 at the beginning of the protocol routing process, c_{ij} is calculated from the b_{ij} using the softmax function.

$$\hat{u}_{j|i} = W_{ij} u_i \quad (5-16)$$

Note: u_i is the output of capsule i in the layer below, and W_{ij} is the weight matrix between each u_i and v_j in the primary capsule, $j \in (1, 3)$.

Use protocol routing between the Primary capsule and the Final CapsLayer. The protocol a_{ij} (formula 5-17) of the update logarithmic probability and coupling coefficient is calculated from formula (5-15).

$$a_{ij} = v_j \hat{u}_{j|i} \quad (5-17)$$

Note: v_j is the vector output of capsule j , and $\hat{u}_{j|i}$ is calculated by formula (5-16).

The final CapsLayer work can be simply summarized in 4 steps :1) Matrix transformation, 2)

Input the weight as shown in formula 5-16, 3) Weighting as shown in formulas 5-14 to 5-15, 4) Non-linear transformation is shown in formula 5-13. Python based pytorch(a deep learning algorithm framework) was used to build CapsNets to complete the experiment [30,31].

3.Results

3.1. Results of STC Method for Screening Saline-alkali Tolerance of Japonica Rice Varieties

According to the saline-alkali tolerance coefficient (STC) of yield, systematic cluster analysis was carried out on 20 japonica rice varieties. The test materials could be divided into three categories at the Euclidean distance of 0.14, as shown in Table 3 and Figure 3.

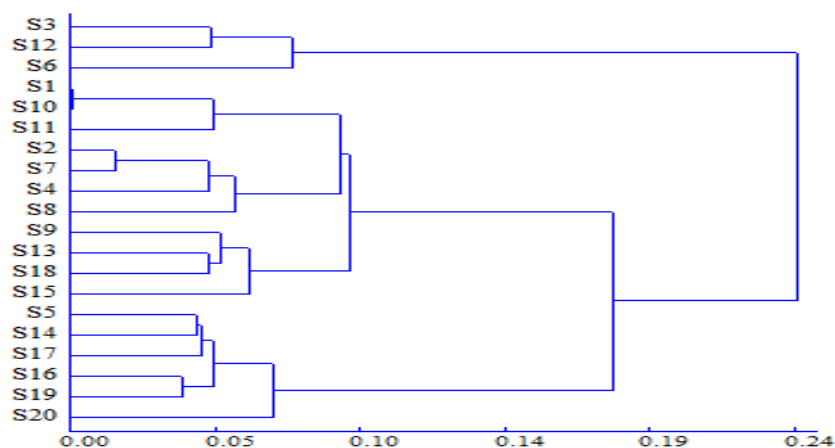


Figure 3. Cluster results of 20 japonica rice varieties tested at 0.14 Euclidean distance

Table 3. Classification results of saline-alkali tolerant japonica rice varieties based on STC

Number	Varieties (lines)	Resistance of salt and alkali	Scope of STC values	Average STC value
S3	SJ6	strong	70.1-100	0.85
S12	QJ28	strong		
S6	LJ21	strong		
S1	KD10	medium		
S10	SJ10	medium		
S11	QJ8	medium	50.1-70	0.64
S2	KD15	medium		
S7	LJ39	medium		
S4	LJ9	medium		
S8	LH05-379	medium		
S9	SJ2	medium	0-50	0.43
S13	DN416	medium		
S18	LQD3	medium		
S15	LD10	medium		
S5	LJ12	sensitivity		
S14	DN9220	sensitivity		
S17	SJ3	sensitivity		
S16	LD12	sensitivity		
S19	JD58	sensitivity		
S20	FY305	sensitivity		

3.2. Results of Raman Spectra of Saline-alkali Tolerant Japonica Rice Varieties

Raman spectra of 20 japonica rice species were obtained using Advantage 532 Raman spectrometer. According to the original two-dimensional image information of the samples (Figure 4), the Raman spectral curves of 661 samples of 20 japonica rice varieties were interleaved, which could not be identified, and various algorithms and models in machine learning were urgently needed for identification.

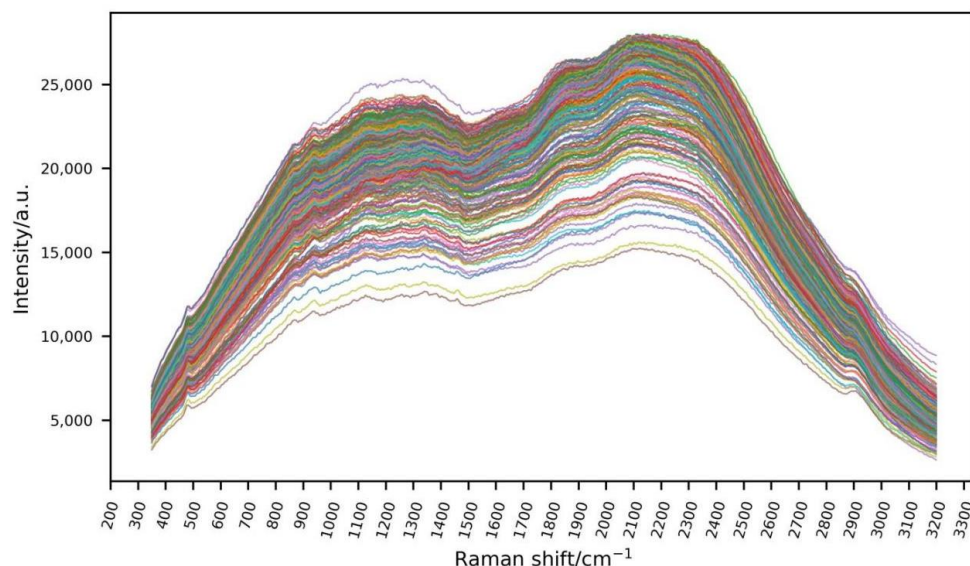


Figure 4. Two-dimensional Raman spectra of 20 japonica rice varieties

3.3. *The Results of Raman Spectral Characteristics of Japonica Rice were Obtained*

After NL data standardization and CWT eliminating baseline drift spectrum smoothing feature extraction, the spectral peak of japonica rice Raman spectrum was obtained, that is, the characteristic information of japonica rice Raman spectrum. It was confirmed by DeltaNu of the United States, the manufacturer of Advantage 532 Raman spectrometer, that 1800cm^{-1} - 2800cm^{-1} was an invalid japonica Raman shift, which had no significance for scientific research. Finally, the 18 spectral peaks are obtained (Figure 5). With reference to previous research results and a large number of relevant literatures [24,25,32], the correlation analysis between spectral peaks of rice Raman spectral characteristics and molecular vibration types was carried out. A clear correspondence between Raman spectral characteristic peaks and molecular vibration types of rice was established, as shown in Table 4.

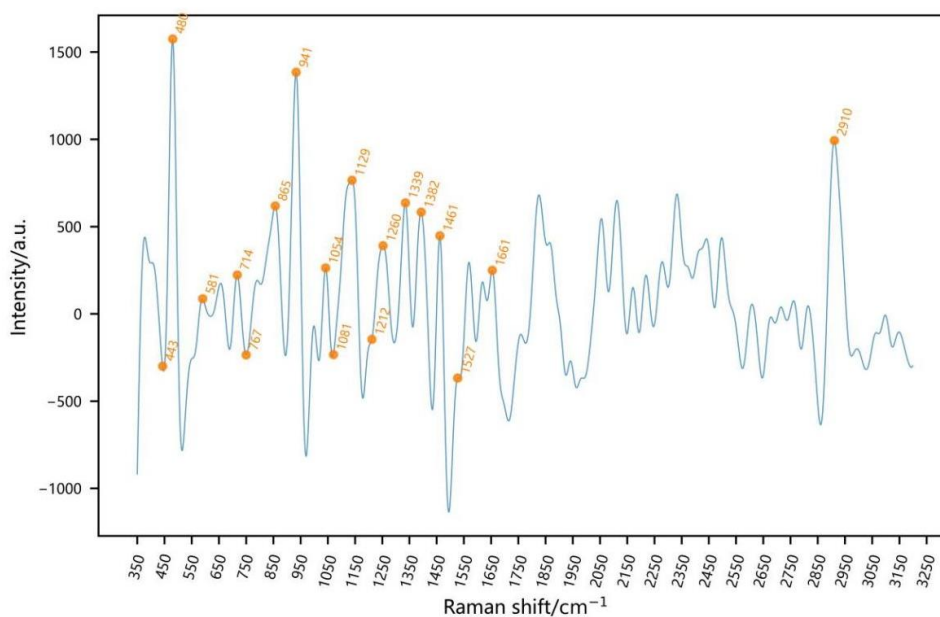


Figure 5. Extraction and visualization of eighteen spectral peaks

Table 4. Molecular vibration types of Raman characteristic spectral peaks in rice

Number	Frequency/cm ⁻¹	Mode of molecular vibration
1	443	pyran ring skeleton vibration
2	480	skeleton vibration peak
3	581	flexural vibration of C-C-O and C-O
4	714	C-C-O flexural vibration
5	767	O=C-N deformation vibration and OH deformation
6	865	CH deformation and C-O ring vibration
7	941	symmetric stretching vibration of C-O-C
8	1054	stretching vibration of C-O, C-C and C-H-O
9	1081	C-O-H bending vibration
10	1129	C-O stretching vibration and C-O-H bending deformation vibration
11	1212	Stretching vibration of C-C and C-O
12	1260	C-N stretching vibration and C-H plane deformation
13	1339	C-C stretching vibration and C-H deformation
14	1382	C-C stretching vibration, CH ₂ and NH ₂ rocking vibration
15	1461	CH plane bending vibration, CH ₂ symmetric vibration and CH ₃ antisymmetric vibration
16	1527	C-N stretching vibration
17	1661	C=O stretching vibration
18	2910	CH ₂ antisymmetric stretching and NH ₂ stretching vibrations

The prominence, width, and dif parameters in SSFP were selected and their combinations were used to obtain the prominence, width, offset, and shape of each spectral peak (see Figure 6). The vertical green line in the figure indicates the prominence and height characteristics of the spectral peak acquired after NL+CWT of japonica rice Raman spectral data. The red horizontal line in the figure indicates the peak width characteristic information of japonica rice Raman spectral data obtained after NL+CWT; The yellow dots in the figure indicate the Raman shift value information of spectral peak obtained after NL+CWT of japonica rice Raman spectral data, that is, the position of peak intensity. The prominence, width, and dif parameters in SSFP were used to jointly obtain the prominence information for each of the 18 peaks in each sample.

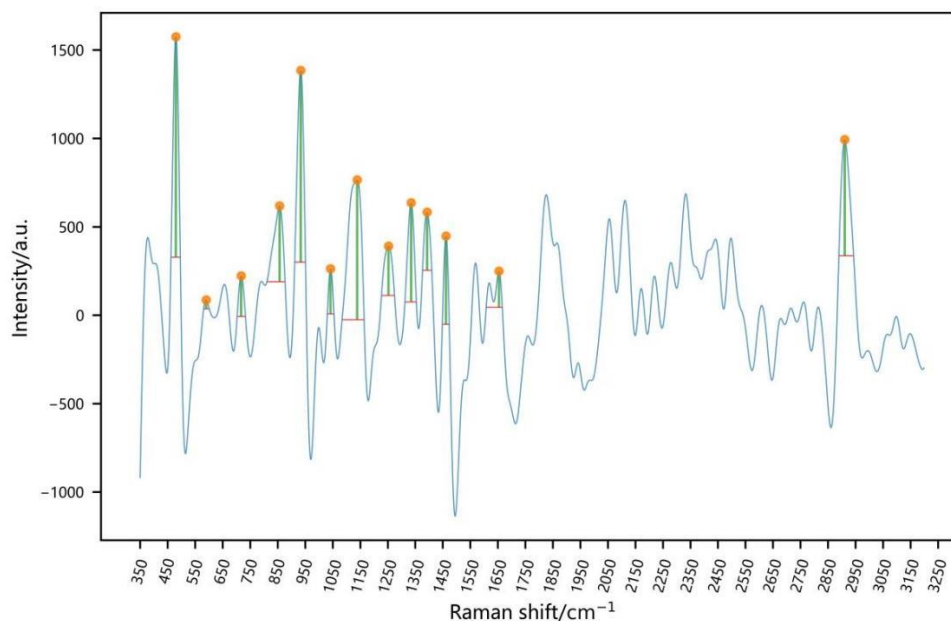


Figure 6. Visual information of three-dimensional characteristics of spectral peak

3.4. Identification Mechanism of Saline-alkali Resistant Japonica Rice by Raman Spectroscopy

3.4.1. Results of Amino Acids Related to Regulation of Saline-alkali Tolerance in Japonica Rice were Obtained

According to authoritative resources such as China Rice Network, 14 genes OsHKT1;5;8[33,34]、OsIAA20[35,36]、OsSIRHC-2[37]、OsMPT3[38]、OsbHLH38[39]、OsMBTB32[40]、OsMADS25[41,42]、OsTPS1[43]、OsSAPK2[44]、RST1[45]、OsSDG721;705[46], which are closely related to saline-alkali tolerance of japonica rice varieties, are from 11 gene families, located on chromosome 1, 4, 5, 6, 7 and 8 of rice, encoding amino acid sequences of different proteins through transcription and translation process. Among them, SKC1 protein encoded by gene OsHKT1;5;8 had 4 amino acids (alanine, histidine, aspartic acid and valine) difference between salt-tolerant rice varieties and salt-sensitive rice varieties. The protein encoded by the other 11 rice saline-alkali tolerance genes was determined by the amount of proline, asparagine and lysine, respectively.

3.4.2. Results of Raman Spectral Characteristic Spectra of Saline-alkali Tolerance Japonica Rice

The following are the structures of 7 amino acids that regulate the salt-alkali tolerance of Raman spectrum of japonica rice (Figure 7) : As shown in screenshot 7 (a) of 3D spatial chemical structure of Proline, the functional group in its side chain (-R) is composed of methylene bond

(-CH) bond and subgroup (-CH₂), corresponding to 865cm⁻¹, 1260cm⁻¹, 1339cm⁻¹, 1461cm⁻¹, 1382cm⁻¹, 1461cm⁻¹ and 2910cm⁻¹ in the Raman spectral characteristic peaks of japonicrice, respectively. As shown in screenshot 7 (b) of 3D spatial chemical structure of Lysine, the functional group in its side chain (-R) is composed of amino group (-NH₂), methylene bond (-CH), subgroup (-CH₂) and amino group (-NH₂), corresponds to 2910cm⁻¹ in the Raman spectral characteristic peaks of japonicrice. As shown in screenshot 7 (c) of the 3D spatial chemical structure of Asparagine, the functional group in its side chain (-R) is composed of amino group (-NH₂), keto group (-C=O), methylene bond (-CH) and subgroup (-CH₂), the keto group (-C=O) corresponds to 1661cm⁻¹ in the Raman spectral characteristic peaks of japonicrice. As shown in screenshot 7 (d) of 3D spatial chemical structure of Alanine, the functional group in its side chain (-R) is composed of methylene bond (-CH) bond and methyl group (-CH₃), and the peak of methyl group (-CH₃) corresponds to the spectral peak at 1461cm⁻¹ in the characteristic spectrum of japonica rice Raman spectrum. As shown in screenshot 7 (e) of 3D spatial chemical structure of Histidine, the functional group in its side chain (-R) is composed of (-CH) bond, subgroup (-CH₂) and imino (-NH). As shown in the screenshot 7 (f) of 3D spatial chemical structure of Aspartate, the functional group in its side chain (-R) is composed of carboxyl group (-COOH), methylene bond (-CH) and subgroup (-CH₂), and the carboxyl group (-COOH) corresponds to the spectral peak at 1661cm⁻¹ in the Raman spectral characteristics of japonica rice. As shown in screenshot 7 (g) of 3D spatial chemical structure of Leucine, the functional group in its side chain (-R) is composed of methylene bond (-CH) bond and methyl group (-CH₃).

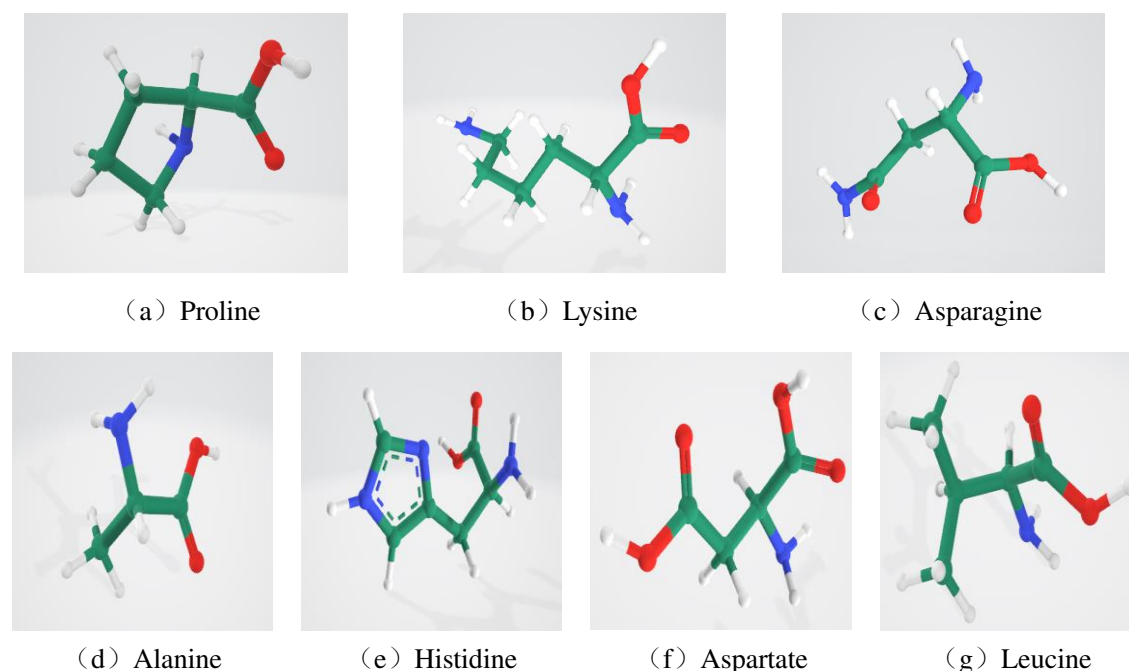


Figure 7. Screenshots of 3D spatial chemical structure of seven different amino acids

By analyzing the Raman scattering effect of rice and combining it with the biological mechanism of salt and alkali tolerance of rice, the Raman spectral salt-alkali tolerance mechanism of rice varieties was analyzed, and the molecular structure and vibration form of amino acid side chain regulating salt and alkali tolerance of japonica rice were confirmed (Table 3). Seven characteristic spectral peaks (865cm⁻¹, 1260cm⁻¹, 1339cm⁻¹, 1382cm⁻¹, 1461cm⁻¹, 1661cm⁻¹ and 2910cm⁻¹) were identified.

3.5. Results of Raman Spectrum Data Preprocessing

As shown in Figure 8, Raman spectrum visualization effect of saline-alkali resistant japonica rice pretreated with NL+CWT+Filtfilt.

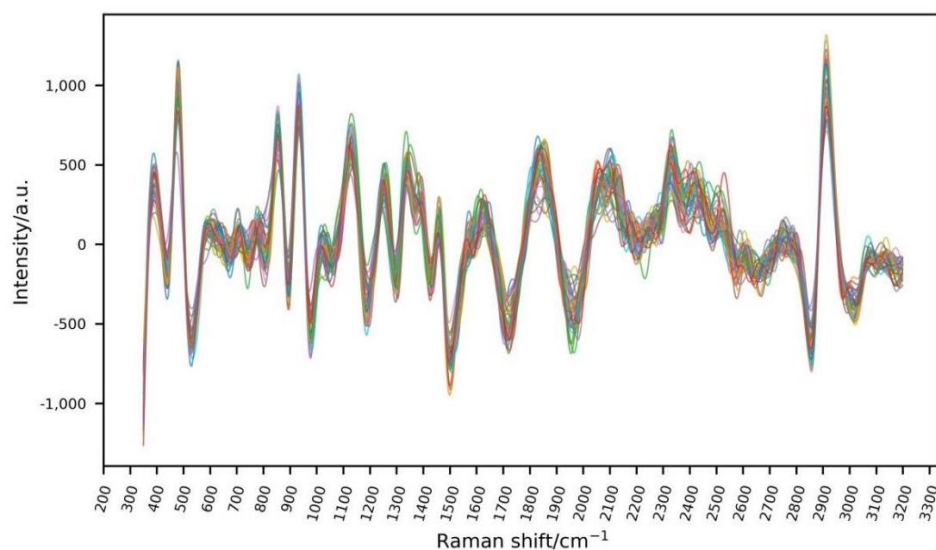


Figure 8. Raman spectra of saline-alkali tolerant japonica rice were pretreated with NL+CWT+Filtfilt

3.6. Identification Results

3.6.1. Result of LR Identification

The identification mechanism of Raman spectra of saline-alkali tolerant japonica rice was studied, and 7 of the 18 characteristic spectra were identified which were closely related to saline-alkali tolerance. Then, all samples were preprocessed. Based on the three-dimensional information of 7 characteristic spectral peaks, LR, which is suitable for linear regression dataset, was used to identify the three-dimensional characteristic information of Raman spectral peaks in the dataset. After the training set and the test set of the data set were divided by 8:2, the LR model was used for identification, and the test accuracy was 72.93% and the time was 0.02 seconds.

3.6.2. Result of SVM Identification

The identification mechanism of Raman spectra of saline-alkali resistant japonica rice was studied, and 7 of the 18 characteristic spectra were identified which were closely related to saline-alkali tolerance. Then all the samples were preprocessed. Based on the 3D information of 7 characteristic spectral peaks, the 3D characteristic information of Raman spectral peaks of the dataset was identified by SVM identification model suitable for discrete data sets. After the training set and test set of the data set were divided according to the ratio of 8:2, the SVM model of four kernel functions was used for identification, and the test accuracy was between 57.89% and 71.43%, and the required time was 0.01s to 27.13s (Table 5).

Table 5. Raman spectrum SVM identification results of saline-alkali tolerant japonica rice varieties based on characteristic spectral peaks

Number	Choice of kernel function	Pretreatment method	Accuracy rate	Time (unit: s)
1	RBF	NL+CWT+Filtfilt	59.40%	0.02
2	LK		71.43%	27.13
3	Simoid		57.89%	0.02
4	PK		66.92%	0.01

3.6.3. Result of CapsNets Identification

Figure 9 shows the 7 spectral peak interceptions of one of the samples on the left. Within the width of 112 Raman shift values and the height of 1600 spectral peak intensity values, the complete spectral peak shape information of each sample can be reflected. In Figure 9, the right is the 865 cm^{-1} spectral peak segmentation diagram of the sample, the blue curve is the morphological feature information of the starting position of the spectral peak, the yellow curve and the green curve are the morphological feature information of the highest point of the rise of the spectral peak and the beginning of the decline of the spectral peak (spectral peak offset), and the red curve is the morphological feature information of the final ending position of the spectral peak. As shown in Figure 10, 28 fragments of information were obtained from each sample, which were sorted according to the order of the starting, highest and ending positions of spectral peaks, and these pictures were fused to obtain a three-dimensional image information for each sample.

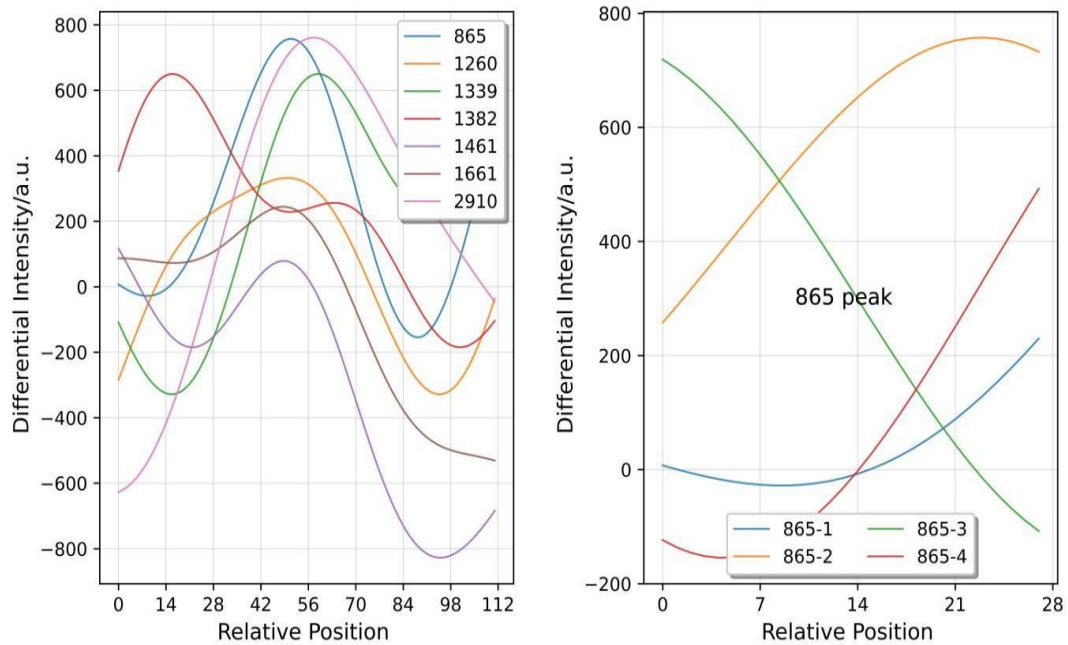


Figure 9. Extracting the shape information of seven characteristic spectral peaks (left) and segmenting the shape information of spectral peaks (right)

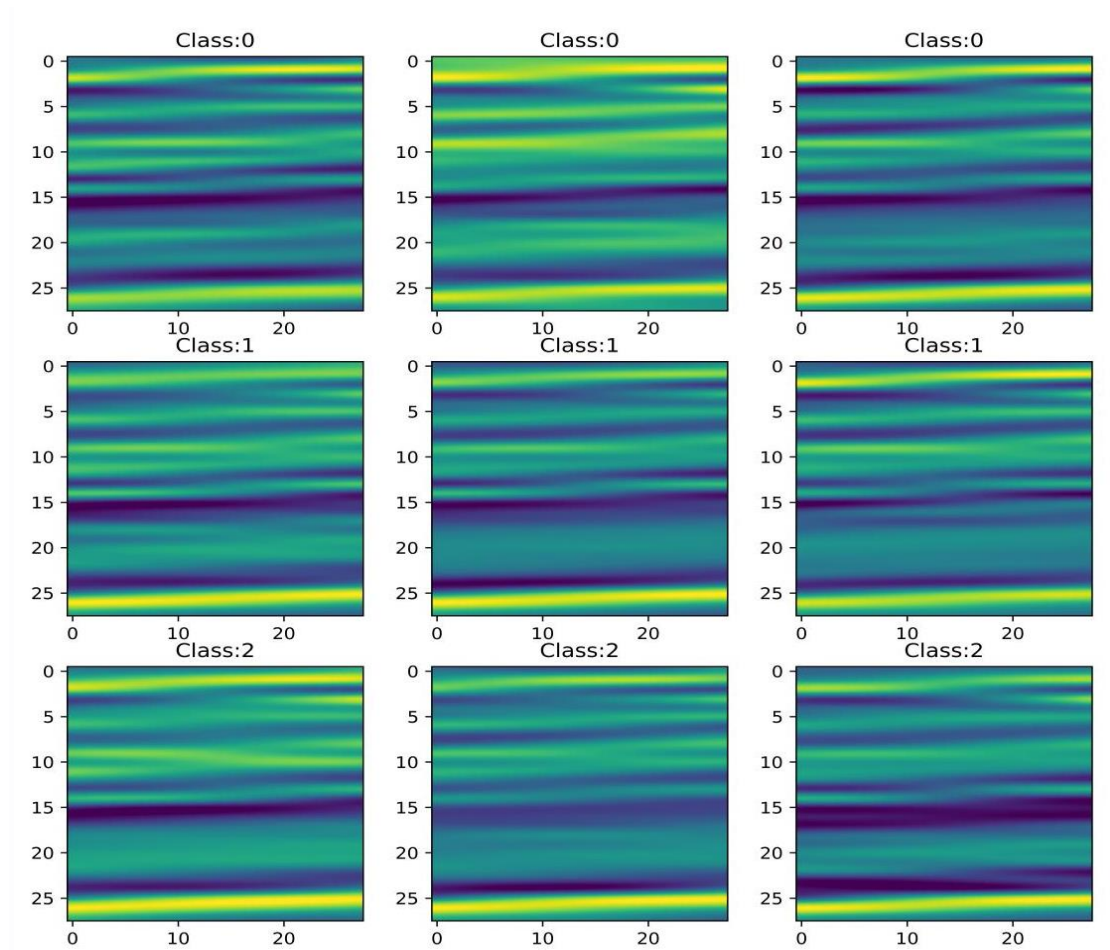


Figure10. Image information obtained by sequentially capturing, segmenting, and fusing samples

Based on the shape characteristics of Raman spectral peaks, 661 2D data sets are transformed into 3D image information by a specific segmentation and fusion method. The 3D image information was divided into a training set and a test set according to the ratio of 8:2, so as to be used for subsequent CapsNets model training and testing. During model training, the Epoch value was set to 300 to ensure that the model had enough iterations to learn and optimize, as shown in Figure 11. When the Epoch value reaches 75, the accuracy of the training set begins to stabilize and reaches the highest value of 95.08%. When the Epoch value reaches 140, the accuracy of the test set also tends to stabilize and reaches the highest value of 80.45%.

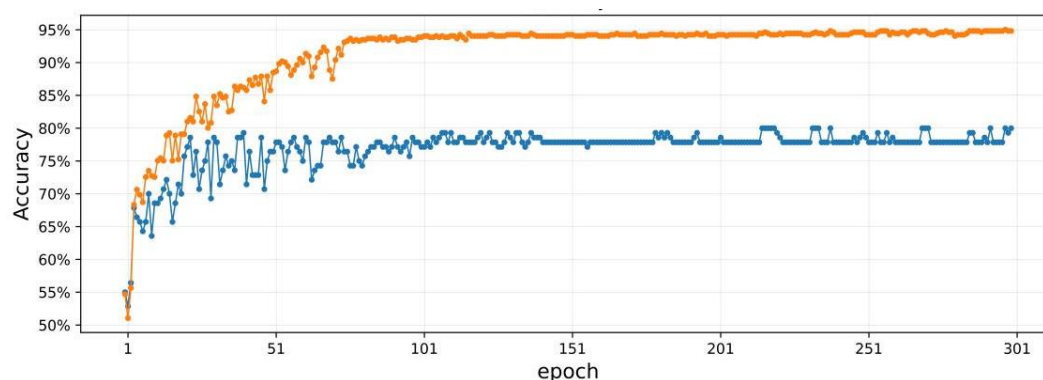


Figure 11. Identification results of CapsNets of saline-alkali tolerant japonica rice varieties with three-dimensional image information

4. Analysis and Discussion

4.1 Raman Spectrum Identification Mechanism of Saline-alkali Resistant Japonica Rice

4.1.1 Analysis of Biological Mechanism of Saline-alkali Tolerant Rice

The molecular vibration types of different Raman spectral peaks of rice are very different, which provides important clues for further study of the characteristics and species of japonica rice. Based on the molecular vibration at the peak of Raman spectrum, the chemical composition and molecular structure characteristics of saline-alkali tolerant rice in different environments can be deeply understood. Saline-alkali tolerant rice refers to rice varieties that can grow and develop normally in high salt and high alkali environments. The tolerance of saline-alkali tolerant rice mainly depends on its gene and protein regulation. In saline-alkali tolerant rice, many genes are considered to be related to saline-alkali tolerance. Some of these genes code for proteins involved in physiological processes such as ion balance, osmoregulation, antioxidant and signal transduction. Amino acids are the building blocks of proteins, which are long chains of several amino acids that are joined together by peptide bonds to form polypeptide chains that further fold into specific three-dimensional structures. Through checking the Chinese rice web, 14 genes were identified from 11 gene families, located on chromosome 1, 4, 5, 6, 7 and 8, encoding amino acid sequences of different proteins through transcription and translation processes. Among them, SKC1 protein encoded by gene OsHKT1;5;8[33,34] had four amino acid differences (Proline, Histidine, Aspartate and Leucine) between salt-tolerant rice varieties and salt-sensitive rice varieties. The protein encoded by the other 11 rice saline-alkali tolerance genes [35-46] was determined by the amount of Proline, Asparagine and Lysine, respectively. The type and order of these amino acid sequences determine the structure and function of the protein, which leads to the difference in the saline-alkali tolerance of rice. Therefore, Proline, Lysine, Aspartate, Leucine, Histidine, Asparagine and Proline play important regulatory roles in saline-alkali tolerant japonica rice, helping japonica rice adapt to saline-alkali stress and maintain normal growth.

4.1.2 Chemical Bond and Vibration Form of Raman Spectral Peaks of Rice

In Raman spectroscopy, spectral peaks represent the scattering of light caused by different modes of vibration. The position, intensity and width of the spectral peaks of different organic compounds are different at the same Raman shift. The difference of spectral peak characteristics depends on the structure of the molecule and the vibration mode of the chemical bond. In order to obtain the spectral peak of japonica rice Raman spectra, CWT, which has the best spectral data standardization and visualization effect, was selected for spectral data baseline elimination to obtain the spectral peak of japonica rice Raman spectra [24,25]. In order to improve the correlation analysis of the obtained 18 japonica rice Raman spectral characteristic peaks, the selection law of Raman spectra and the types of molecular vibration, and to conduct in-depth research on the results combined with a large number of relevant literatures (Table 4), it was found that the following rules were followed:

- (1) The types of molecular vibration at the obtained 18 spectral peaks are different.
- (2) The vibration bond of the same molecule and the vibration type are different, such as the CO molecule at number 6 is the ring vibration of C-O single bond, and the CO molecule at number 17 is the C=O double bond stretching vibration.
- (3) The vibration types of the same molecules are different, such as C-O-H at number 9 is bending vibration, and C-O-H at number 10 is bending deformation vibration.

(4) The vibration of the same molecule is located at different peak positions, the CH molecular vibration is located at serial numbers 6, 12, 13 and 15, the CH₂ molecular vibration is located at serial numbers 14, 15 and 18, and the CC molecular vibration is located at serial numbers 8, 13 and 14.

4.1.3. Identification of main characteristic spectral peaks in saline-alkali resistant japonica rice

By analyzing the molecular chemical structures and side chain chemical bonds of the above 7 amino acids (Figure 7), it can be found that amino acid side chains of saline-alkali tolerant japonica rice contain imino (-NH), amino group (-NH₂), keto group (-C=O), methylene bond (-CH) bond, subgroup (-CH₂), methyl group (-CH₃) and carboxyl group (-COOH). By analyzing the Raman scattering effect of rice and combining it with the biological mechanism of salt and alkali tolerance of rice, the mechanism of salt and alkali tolerance of rice varieties by Raman spectrum was analyzed, and the molecular structure and vibration form of amino acid side chain regulating salt and alkali tolerance of japonica rice was confirmed. Seven characteristic spectral peaks (865cm⁻¹, 1260cm⁻¹, 1339cm⁻¹, 1382cm⁻¹, 1461cm⁻¹, 1661cm⁻¹ and 2910cm⁻¹) were identified.

4.2. Analysis of Raman Spectral Identification Method for Saline-alkali Tolerance Japonica Rice Varieties based on Characteristic Spectral Peaks

4.2.1. LR and SVM Identification Models

When the features of the data set are linearly separable, the LR model can accurately identify a given set or multiple sets of data [47,48] and perform well. Therefore, the Raman spectrum identification of saline-alkali tolerant japonica rice varieties was carried out based on the three-dimensional characteristic information of the characteristic spectral peak as the input of LR identification model, and the accuracy of 72.93% was obtained. When the features of the data set are linearly indivisible, SVM can map the low-dimensional data into the high-dimensional space through the kernel function to make it linearly divisible [49]. Therefore, the Raman spectrum identification of saline-alkali tolerant japonica rice varieties was carried out based on the three-dimensional characteristic information of the characteristic spectral peak as the input of SVM identification model, and the accuracy of 71.43% was obtained. By selecting the traditional identification model LR and SVM, the difference in accuracy of the three classifications obtained is only 1.5%, and the characteristics of the experimental data set are determined to be linearly separable and linearly indivisible, that is, mixed type.

4.2.2. CapsNets Identification Model

Aiming at the current Raman spectral data set of non-large-scale salt-alkaline tolerance japonica rice varieties, CapsNets requires the model to learn feature variables in the capsule to maximize the retention of those valuable feature information, so that possible variables can be inferred with less training data [27, 50]. As shown in Figure 11, when the Epoch value reaches 75, the accuracy of the training set begins to stabilize and reaches the highest value of 95.08%, which indicates that the model has strong learning ability on the training data and can effectively extract and recognize the feature information in three-dimensional images. When the Epoch value is 140, the accuracy of the test set also tends to be stable and reaches the highest value of 80.45%. Although the accuracy of the test set is slightly lower than that of the training set, considering that the test set is used to evaluate the model's performance on unknown data classification attributes,

this result still indicates that the model has good generalization ability.

4.2.3 Conversion of 2D Data Information to 3D Image Information

When the information presented by different samples in two-dimensional space is so similar that it is difficult to distinguish them (Figure 4), converting them to three-dimensional space may reveal differences that would otherwise be hidden in two-dimensional space [27]. Therefore, according to the characteristics of Raman spectral information of saline-alkali tolerant japonica rice varieties, the shape characteristics of 7 selected spectral peaks were converted from 2D data information to 3D image information by means of one intercept, one segmentation and one fusion. As shown in Figure 10, each picture represents the result of converting a sample from two-dimensional data information to three-dimensional image information. The first is three-dimensional image information of three randomly selected saline-alkali sensitive japonica rice varieties, the second is three-dimensional image information of three randomly selected saline-alkali tolerant japonica rice varieties, the third behavior is three-dimensional image information of 3 randomly selected japonica rice varieties with strong saline-alkali tolerance. From the comparison of 3D image information, the difference of 3D image information of different varieties of the same saline-alkali tolerance category was not obvious, the 3D image information of different saline-alkali tolerance categories was significantly different, especially the 3D image information of saline-alkali sensitive japonica rice samples was significantly different from that of strong saline-alkali tolerant japonica rice samples. Through the conversion of two-dimensional data information to three-dimensional image information, the identification model of CapsNets was constructed and showed the best identification results, which could be regarded as the optimal method in the identification of saline-alkali tolerance japonica rice varieties.

Author Contributions: B.M. and D.M. conceived the study and designed the project. R.L. and X.Z. performed the experiment, analyzed the data, and drafted the manuscript. B.M., H.X., J.H., K.T., M.Y. and L.W. helped to revise the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Scientific Research Project of Heilongjiang Provincial Scientific Research Institutes(CZKYF2023-1-B013), Heilongjiang Province Key Research and Development Plan Project (JD2023GJ04) and Key project of science and technology plan of Qiqihar City(ZDGG-202307).

Acknowledgments: I hereby thank Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences in China for providing me with japonica rice materials, Heilongjiang Bayi Agricultural University in China for providing Raman spectroscopy and other instruments.

Data Availability: The data and code presented in this study are openly available at:

<https://github.com/mabo8210/Mechanism-of-Saline-alkali-Tolerance>.

Conflicts of Interest: The authors hereby declare that there are no conflict of interest in the present study.

Reference

- [1] Liu anjin, Jiang Yu, Shang Quanyu, et al. Research Progress on Improvement Technology of Rice-Alkali Land in Heilongjiang Province[J].Agricultural Science in Heilongjiang Province,2022,(08):83-86+90.
- [2] Ma Bo, Liu Chuanzeng, Hu Jifang, Liu Kai, Zhao Fuyang, Wang Junqiang, Zhao

- Xin, Guo Zhenhua, Song Lijuan, Lai Yongcai, Tan Kefei. Intelligent Identification and Features Attribution of Saline–Alkali-Tolerant Rice Varieties Based on Raman Spectroscopy[J]. *Plants*, 2022, 11(9).
- [3] Zhengjun Wang, Jingjing Zhuang, Anping Zhao, Xinxin Li. Types, harms and improvement of saline soil in Songnen Plain[J]. *IOP Conference Series: Materials Science and Engineering*, 2018, 322(5).
- [4] Sun Yu, Wang Qi, Song qiulai, Song qiulai, et al. Research Progress on Response of Rice to Saline-alkali Stress and Physiological Mechanism of Saline-alkali Tolerance in Cold Region[J]. *Agricultural Science in Heilongjiang Province*, 2022(08):57-61.
- [5] Hang Xiao, Ma Wendong, Wang Guiling, et al. Research Progress on Physiological and Genetic Mechanism of Saline-Alkali Tolerance in Rice[J]. *Agricultural Science in Heilongjiang Province*, 2022(08):62-67.
- [6] Li Hongyu, Li Yi, et al. Principal Component Analysis and Comprehensive Evaluation of Saline-Alkaline Tolerance Related Traits of Northern Japonica Rice[J]. *Journal of Nuclear Agronomy*, 2020,34(08):1862-1871.
- [7] Li Yi. Principal Component Analysis and Comprehensive Evaluation of Saline-alkaline Tolerance Related Traits in Rice in Cold Regions[D]. Heilongjiang Bayi Agricultural University, 2022.
- [8] Wang Junhe, Liu Chuanzeng, Ma Bo, et al. Breeding and High-yield Cultivation Techniques of High-quality Fragrant Rice Qijing 10 in Cold Region[J]. *Chinese seed industry*, 2019(09):80-82.
- [9] Shaghaleh H, Azhar M, Rehman U Z M , et al. Effects of agro based organic amendments on growth and cadmium uptake in wheat and rice crops irrigated with raw city effluents: Three years field study[J]. *Environmental Pollution*, 2024,344123365-.
- [10] Wang Zhishan, Zhu Yongqiang, Li Ni, Liu Hai, Zheng Huajun, Wang Weiping, Liu Yang. High-throughput sequencing-based analysis of the composition and diversity of endophytic bacterial community in seeds of saline-alkali tolerant rice[J]. *Microbiological Research*, 2021, 250.

- [11] S.Geetha, A. Vasuki, P. JagadeeshSelvam, R. Saraswathi, S. L. Krishnamurthy. Palanichamy, Manikandan,M. Dhasarathan,G. Thamodharan, M. Baskar. Develop ment of sodicity tolerant rice varieties through marker assisted backcross breeding[J]. Electronic Journal of Plant Breeding, 2017, 8(4).
- [12] Yue S, Zhongkai C, Huizhen C, et al. Analysis of the Genetic Stability of Insect and Herbicide Resistance Genes in Transgenic Rice Lines: A Laboratory and Field Experiment. [J]. Rice (New York, N.Y.),2023,16(1):8-8.
- [13] Hibben J. H., Teller E. The Raman effect and its chemical aplications and physical research. Industrial and Engineering Chemistry. News Ed. 1939, 17, 556.
- [14] Herzberg G. Molecular spectra and molecular structure. In Infrared and Raman Spectra of Polyatomid Molecules; Van Nostrand, R., Ed.; American Journal of Physics: New York, NY, USA, 1945; Volume 2.
- [15] Shipp D.W., Sinjab F., Nothingher I. Raman spectroscopy: Techniques and applications in the life sciences. Adv. Opt. Photonics 2017, 9, 315–428.
- [16] Alian W., Randy L.K., Bradley L., Jolliff Z.L. Raman imaging of extraterrestrial materials. Planet. Space Sci. 2015, 112, 23–34.
- [17] Vankeirsbilck T., Vercauteren A., Baeyens W., Van D.W.D. Applications of Raman spectroscopy in pharmaceutical analysis. Trends Aanlytical. Chem. 2002, 21, 869–877.
- [18] Kong H. P., Liu Z. Q. Implementation of Dimension Reduction Method of On-board Data of High-speed EMU Based on Python. Comput. Eng. Softw. 2020, 41, 114–117.
- [19] Perkel J. M. Python power-up: New image tool visualizes complex data. Nature 2021, 600, 347–348.
- [20] Cao S., Zeng Y., Yang S., Cao S. Research on Python Data Visualization Technology. J. Phys. Conf. Ser. 2021, 1757, 012122.
- [21] Martin L., Bramkamp Y., Köster T., Staiger D. SEQing: Web-based visualization of iCLIP and RNA-seq data in an interactive python framework. BMC Bioinform. 2020, 21, 113.

- [22] Liu Rui, Tan Feng, Wang Yaxuan, Ma Bo, Yuan Ming, Wang Lianxia, Zhao Xin. Machine Learning Identification of Saline-Alkali-Tolerant Japonica Rice Varieties Based on Raman Spectroscopy and Python Visual Analysis[J]. Agriculture, 2022, 12(7).
- [23] Ma Bo. Plant Immunity Inducers Induced Rice Resistance to Rice Seedling Blight and its Mechanism[D].Northeast Agricultural University,2020.
- [24] Wang Yaxuan. Study on mechanism and system of rice origin Identification based on Raman spectroscopy[D].Heilongjiang Bayi Agricultural University,2022.
- [25] Tian Fangming. Identification of Rice Based on Analysis of Raman Spectrum and Organic Ingredients[D].Jilin University,2018.
- [26] Zhang F., Wang D. Vibration Signal Filtering Algorithm Based on Singular Value Subspace Decomposition[J]. Revista Tecnica De La Facultad De Ingenieria Universidad Del Zulia, 2017.
- [27] Kumbhar D., Palliyarayil A., Reghu D., Shrungar D., Umapathy S., Sil S. Rapid discrimination of porous bio-carbon derived from nitrogen rich biomass using Raman spectroscopy and artificial intelligence methods[J]. Carbon, 2021, 178.
- [28] Tang L Y, Li N Y, Dong X, et al. Study on COD Emission Reductions Assignment of Basin Based on Information Entropy Method and Gini Coefficient Method[J]. Advanced Materials Research, 2014, 3400(1014-1014).
- [29] Vladimir N, Michiel K. On Stochastic Optimization and Statistical Learning in Reproducing Kernel Hilbert Spaces by Support Vector Machines(SVM). Informatica 2009, 20, 273–292.
- [30] Liu Xinghan, Huo Hua. Automatic summarization for text based on mutual information[J]. Journal of Hefei University of Technology (Natural Science Edition),2014, 37(10):1198-1203.
- [31] O'Shea K, Nash R. An Introduction to Convolutional Neural Networks[J]. CoRR, 2015, abs/ 1511. 08458.
- [32] Yaxuan W, Feng T. Extraction and classification of origin characteristic peaks from rice Raman spectra by principal component analysis[J]. Vibrational

Spectroscopy,2021,114.

- [33] FELIX HAUSER, TOMOAKI HORIE. A conserved primary salt tolerance mechanism mediated by HKT transporters:a mechanism for sodium exclusion and maintenance of high K⁺/Na⁺ ratio in leaves during salinity stress Plant, Cell & Environment, 2010, 33(4): 552-565.
- [34] Blanca Garciadeblás, María E. Senn; María A. Bañuelos; Alonso Rodríguez-Navarro Sodium transport and HKT transporters:the rice model The Plant Journal, 2003, 34(6): 788-801.
- [35] Aiyuan Zhang, Xu Yang, Jia Lu, Fangyuan Song, Jinghuan Sun, Cong Wang, Juan Lian, Lili Zhao, Baocun Zhao. OsIAA20,an Aux/IAA protein,mediates abiotic stress tolerance in rice through an ABA pathway Plant Science. 2021, 308: 110903.
- [36] Yaling Song, Lei Wang, Lizhong Xiong Comprehensive expression profiling analysis of OsIAA gene family in developmental processes and in response to phytohormone and stress treatments Planta. 2009, 229(3): 577-591.
- [37] Min Seok Choi. Cheol Seong Jang Overexpression of Rice C3HC4-Type RING Finger Protein Gene, OsSIRHC-2, Improves Salinity Tolerance Through Low Na⁺ Accumulation Journal of Plant Biology. 2023, 66(2): 147-16.
- [38] Shengcai Huang, Shichao Xin, Guoqiang Xie, Jiao Han, Zhonglai Liu, Bing Wang, Shuqing Zhang, Qingyu Wu, Xianguo Cheng Mutagenesis reveals that the rice OsMPT3 gene is an important osmotic regulatory factor. The Crop Journal, 2020, 8(3): 465-479.
- [39] Fengping Du, Yinxiao Wang, Juan Wang, Yingbo Li, Yue Zhang, Xiuqin Zhao, Jianlong Xu, Zhikang Li, Tianyong Zhao, Wensheng Wang, Binying Fu. The basic helix-loop-helix transcription factor gene,OsbHLH38, plays a key role in controlling rice salt tolerance Journal of Integrative Plant Biology, 2023.
- [40] Uzair Ullah, Wenli Mao, Waseem Abbas, Badr Alharthi, Nadeem Bhanbhro, Meng Xiong, Nazish Gul, Abdullah Shalmani. OsMBTB32, a MATH-BTB domain-containing protein that interacts with OsCUL1s to regulate salt tolerance in rice Functional & Integrative Genomics, 2023, 23(2):139.

- [41] Chunyan Yu, Yihua Liu, Aidong Zhang, Sha Su, An Yan, Linli Huang;Imran Ali, Yu Liu, Brian G. Forde;Yinbo Gan MADS-box Transcription Factor OsMADS25 Regulates Root Development through Affection of Nitrate Accumulation in Rice PLoS ONE, 2015, 10(8): e0135196.
- [42] Ning Xu, Yanli Chu, Hongli Chen, Xingxing Li, Qi Wu, Liang Jin, Guixue Wang, Junli Huang. Rice transcription factor OsMADS25 modulates root growth and confers salinity tolerance via the ABA-mediated regulatory pathway and ROS scavenging PLoS Genetics, 2018, 14(10): e1007662.
- [43] Hao-Wen Li, Bai-Sheng Zang, Xing-Wang Deng, Xi-Ping Wang. Overexpression of the trehalose-6-phosphate synthase gene OsTPS1 enhances abiotic stress tolerance in rice Planta, 2011, 234(5): 1007-1018.
- [44] Dengji Lou, Houping Wang, Diqu Yu. The sucrose non-fermenting-1-related protein kinases SAPK1 and SAPK2 function collaboratively as positive regulators of salt stress tolerance in rice BMC Plant Biology, 2018, 18: 203.
- [45] Ping Deng, Wen Jing, Chengjuan Cao, et al. Transcriptional repressor RST1 controls salt tolerance and grain yield in rice by regulating gene expression of asparagine synthetase Proceedings of the National Academy of Sciences of the United States of America, 2022, 119(50): e2210338119.
- [46] Pengfei Jiang, Shiliang Wang, Aziz Ul Ikram, Zuntao Xu, Haiyang Jiang, Beijiu Cheng, Yong Ding. SDG721 and SDG705 are required for rice growth Journal of Integrative Plant Biology, 2018, 60(7): 530-535.
- [47] http://www.360doc.com/content/21/0315/20/68068867_967146874.shtml
- [48] <https://zhuanlan.zhihu.com/p/205464132>
- [49] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.calinski_harabasz_score.html#sklearn.metrics.calinski_harabasz_score
- [50] Nattino G, Pennell M L, Lemeshow. S. Assessing the goodness of fit of logistic regression models in large samples: A modification of the Hosmer-Lemeshow test. Biometrics 2020, 76, 549–560.